Information Recommendation System Based on the Analysis of User Relationship and Micro- blogging Content Mining

Fawang Han

Nanjing Forest Police College, Jiangsu Nanjing 210023, China
hanfa1wang@163.com

Based on analysis of user relationship to create a user influence model. By means of collecting the user data of Sina Weibo, this thesis will analyse Weibo users relationship and the causes of customer relationship networks. By using the fuzzy comprehensive evaluation methods to determine the relationship between different factors that may affect user influence and propose a formula as well as construct the centre user's user influence model. Proposed the recommendation algorithm of micro-blogging information. In this paper, the existing information recommendation algorithms were reviewed and summarized some micro-blogging information recommended methods. On account of this, the thesis proposes another calculation method which is based on the user relationship and micro-blogging content analysis to make the Weibo information recommendation for centre audience, and verifies the effectiveness of this algorithm through the experimental results.

1. Introduction

Sina Weibo is a customer-centric platform whose success is ultimately judged by its users. Currently, the registered users of Sina Microblog are over 300 million, making it a large networking groups on site. In this thesis, the users as the primary data source of customer relationship analysis as well as the primary audience of this recommendation services, are decisive factors to build user influence model.

In the world of Weibo, each user has three important attributes - "followed", "follower" and "post number". This three items, to some extent, explained the type of a user (Bai and Lin, 2015). The number of "followed" indicates the breadth of information that he covers; The number of "followers" shows the user's popularity degree in the microblog groups; And the" post number "is the best embodiment of the user's activity level. By the orders of magnitude of these data, we can get two different categories of users, which are ordinary users and authenticated users (Guo, 2013). Authenticated user is a celebrity authentication system launched to protect the interests of celebrities. The mark of celebrity certification is a "V" sign followed by the name of an authenticated user. The certified celebrity including microblog of programs (such as "The Voice of China"), media microblog (such as "Zhejiang Satellite TV"), microblog of popular celebrities (such as "Yao Chen"), government or agency's official microblog (such as "Chengdu issued") and enterprise microblog (such as "Lenovo"), etc(Jia, 2012). Currently, an authenticated Sina user is more than 300,000, of which there are over 130,000 corporation and institution accounts. These users and ordinary users may also follow each other which is very important for constructing the model of user influence (Hwang and Choi, 2016).

2. Experimental data

This study is based on the largest domestic microblog platform - Sina Weibo which provides its users with API that can capture microblog. But most of the API's visit, for example to visit published microblogs, to obtain private message followers, all need user authentication(Yu and He, 2016). This thesis, based on the user relationship information recommendation, through the microblog read interface, mainly captures the data of user relationship and related microblog posts, which is shown in details shown in Figure 1.

The capture of microblog post is based on user information watch-list to capture out layer by layer in a divergent manner. After the observation and study of the users on microblog platform, it can be found that Weibo users have different relationship network compared with other social platforms. With the purpose of this

paper, the author depicts a shoot outward user-centric relationship network (Li, 2016), which can be succinctly expressed in Figure 2. User1 is a centre user, while user2, user3 and user4 are a collection of user group, called related users composed by authenticated users. It is noteworthy that for an ordinary user as a central user, the number of authenticated user is much larger than ordinary user in User 2 Group, for example, the followers of celebrities. While in user 3 Group, the average user is likely the dominant, that is to say, there may be related friends and classmates in real life. As for user 4 Group, the authenticated user is likely to be 0, and large number of ordinary users are "Zombies", some fake fans that can be bought with money(Zhou et al, 2013).

Based on above analysis, when building user influence model, firstly we should analyse the position of related users in the centre user outwardly diverging network, and assign weights for them (Yuan and Hu, 2013).

To recommend some concerned Weibo for centre users, and from the aspect of user relationship mining, we mainly deal with centre users' concerns instead of processing their fans(not including mutual powder). The author believes that if a centre user wants to get information of related users, he need to follow them directly or indirectly. As for the followers, since centre user choose not to follow them, the related information of them has no influence and does not be cared by the centre user.
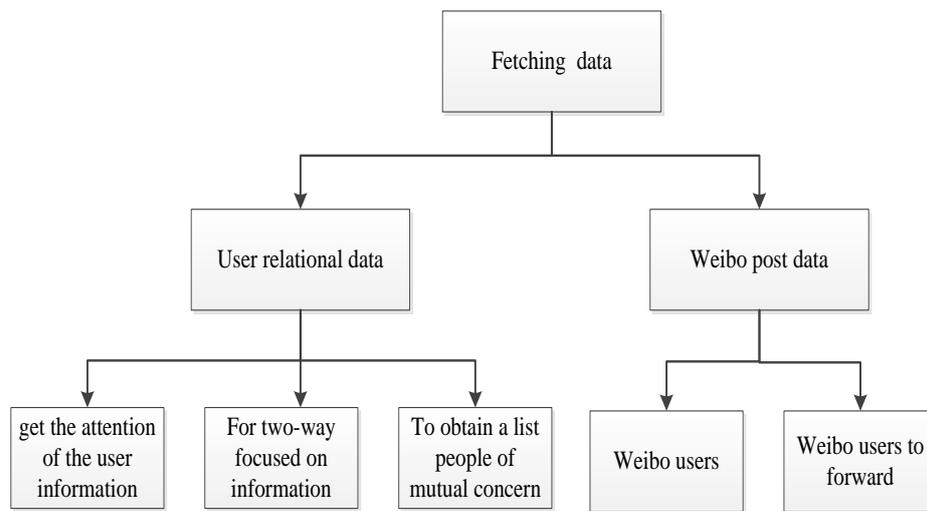

Figure 1: Weibo information fetching

This study uses the API interfaces of Sina M to take out two layers users of a centre user, that is to say, what we captured is the related users of a centre user and their related users. This diagram is shown in Figure 3.

In order to calculate an ordinary user's influence toward centre user, we need to give each user in Figure 3 the weight value and to increase the weight of some users since their impact on the centre user may be increased according to his position in the customer network.

For example, in Figure 3, the related user not only is followed by centre user1, but also got followed by relevant user 2. Therefore, we enhance the relevant user1's weight since it directly and also indirectly be followed by centre users. We provide different weight values with users, and take the centre user 1215155501 of Sina microblog as an example to list the number of users under different weight value.

Account of above analysis, this thesis identifies four major factors that influence the construction of user influence model

(1) The position of related users on the user relation network:

We have analysed this part and provided corresponding weight value of related users.

(2) Whether related users are authenticated users

From above analysis, it is found that a celebrity has a great influence in the Weibo users influence study. So we put more attention on the "Verified" users in collecting data. For example, an actress user's posts are likely to have a greater impact on their groupies.

(3) Whether centre users and their posts are mentioned or re-tweeted by related users:

Needless to say, when a centre user is @ by related users, it is more likely that the related users provide some information that the centre user may interested in. Now a lot of researches discuss the willingness of users to forward information, while trust is a very important study variable in the information system. Based on the long-term relationship with users, the performance and benefits of information system, the users'

behaviour willingness and loyalty will all change. It is the trust between the centre user and the related user that makes the recommendation information relevant user become valuable.
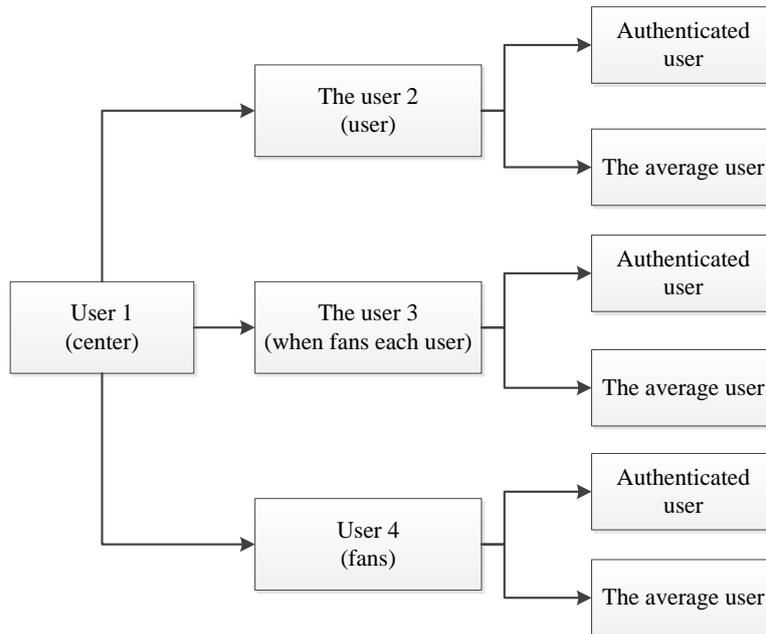

Figure 2: User network

3. Personalized information recommendation method

The arrival of Internet era makes large demand of users being satisfied. However, with the substantial growth of overall information, users cannot find the content they need from overload information are. Up to now, the only solution of this is to establish an information access system which can be quickly retrieve information through retrieval system, but still cannot meet the individual needs. So the personalized recommender system came into being. This system combines information filtering technology and decision support technology to solve the problem of overloading information. On the basis of users 'behaviour and preferences, it makes recommendation with similar contents, so that the right information is presented to different users. The universal model of personalized recommendation system is shown in Figure 5:

Among them, the recommendation algorithm module is the core of overall system, whose performance determines the success of entire information recommendation system. Currently, there are many recommendation algorithms, including: content-based recommendation, collaborative filtering, social networks-based recommendation and other recommendations based on a combination of recommendation and network structures.

The most successful recommendation algorithm existing now is the collaborative filtering recommendation algorithm which considered people's daily life habits, and divide into three models: projects collaborative recommendation, user's collaborative and filtering collaborative recommendation. The core of users collaborative recommendation is neighbours queries that can find the similarity between users, and the higher the similarity, the more similar the users are, then come to the information recommendation. Similar to users collaborative recommendation, projects collaborative recommendation is also based on the users' trust to a brand. Put it another way, if more people believe in on brand, there will be many users choose the brand of products. While recommendation based on model collaborative is somewhat different.

The difference is the machine learning and data applications on existing statistical methods for the testing model. Collaborative filtering algorithms have obvious advantages, which can be used for complex unstructured objects, help users discover new points of interest, and will gradually improve its performance with the increasing number of users. The disadvantage is that when the number of users increases, polyethylene and what they interested in are not the same, the comment of users may be different, and it cannot recommend the new user information. In addition, historical data has a great influence on the recommendation quality.

The recommended method that based on the network structure is a relatively new study which makes the user and recommend objects to be the nodes without considering the contents of the user and the recommended

objects. Once the user selects an object, the algorithm will automatically think that information is hidden in the relations among them. The advantage of this system is to recommend some unpopular items to its users, which is believed by the provider that recommendation on unpopular items will reflect the effect of the recommendation system better since popular items are basically users do not need extra recommendation work .Due to a variety of recommendation algorithm has its own advantages and disadvantages, so in practical applications, we can combined different approaches to solve the problem by complement each other and against weaknesses. There are two main ideas of mixing two recommends, mix of recommendation result, and the other is combined with recommendation algorithm. The specific application should be based on practical problems.
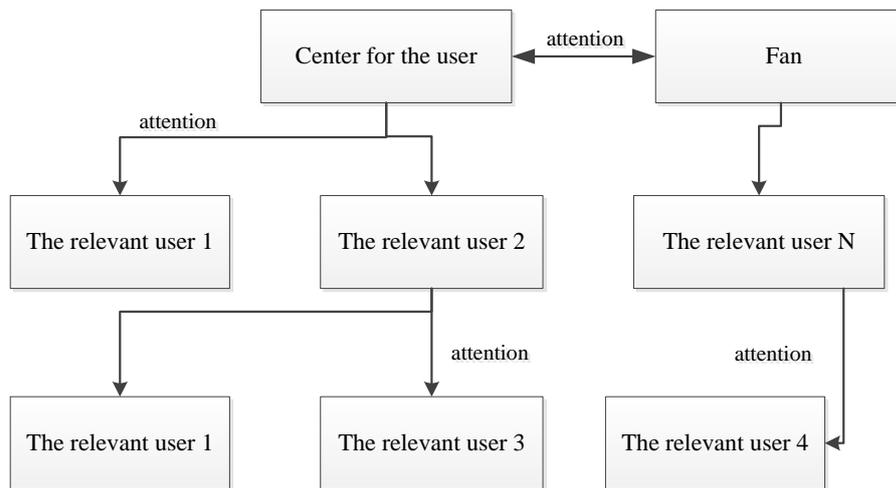

Figure 3: Focus on centre users

4. Microblogging recommended method

So far, tweets recommendation still belongs to a relatively new research, algorithm proposed by researchers is basically based on the traditional information recommendation algorithm, and generally can be divided into two categories: neighbouring users analysis and matrix decomposition. A common application of Users neighbourhood analysis is based on user tags tweets recommendation. Through analysis of the users characteristics, making classification and collection of user labels and tags, and by analysing the similarity of users characteristics and weight tweets recommendation, the effect of this recommendation algorithm is better, but requires a large amount of tag data support which is not a very reliable solution for dynamic user.

In terms of a matrix decomposition proposed by Wang Sheng [74] Bayesian personalized Weibo recommendation algorithm is to compare the new algorithm, the algorithm takes this context microblog and time factor into account to solve recommended problem with the low level of activity, this method caress less dynamic microblog and user interest issues. Meanwhile, the combination of matrix decomposition model in their study is based on the use of microblog to solve the data sparseness and asymmetry. However, this algorithm ignores the relationship between users and microblog system in the cold start phenomena which need to be further improved.

Because microblog convenience, effectiveness, and user dynamics, etc., tweets recommendation algorithm in data mining is still a very important and urgent subject.

The most successful recommendation algorithm existing now is the collaborative filtering recommendation algorithm which considered people's daily life habits, and divide into three models: projects collaborative recommendation, user's collaborative and filtering collaborative recommendation.

The core of users collaborative recommendation is neighbours queries that can find the similarity between users, and the higher the similarity, the more similar the users are, then come to the information recommendation. Similar to users collaborative recommendation, projects collaborative recommendation is also based on the users' trust to a brand. Put it another way, if more people believe in on brand, there will be many users choose the brand of products. While recommendation based on model collaborative is somewhat different.

The difference is the machine learning and data applications on existing statistical methods for the testing model. Collaborative filtering algorithms have obvious advantages, which can be used for complex unstructured objects, help users discover new points of interest, and will gradually improve its performance

with the increasing number of users. The disadvantage is that when the number of users increases, polyethylene and what they interested in are not the same, the comment of users may be different, and it cannot recommend the new user information. In addition, historical data has a great influence on the recommendation quality.

The recommended method that based on the network structure is a relatively new study which makes the user and recommends objects to be the nodes without considering the contents of the user and the recommended objects. Once the user selects an object, the algorithm will automatically think that information is hidden in the relations among them.

The advantage of this system is to recommend some unpopular items to its users, which is believed by the provider that recommendation on unpopular items will reflect the effect of the recommendation system better since popular items are basically users do not need extra recommendation work .Due to a variety of recommendation algorithm has its own advantages and disadvantages, so in practical applications, we can combined different approaches to solve the problem by complement each other and against weaknesses. There are two main ideas of mixing two recommends, mix of recommendation result, and the other is combined with recommendation algorithm. The specific application should be based on practical problems.
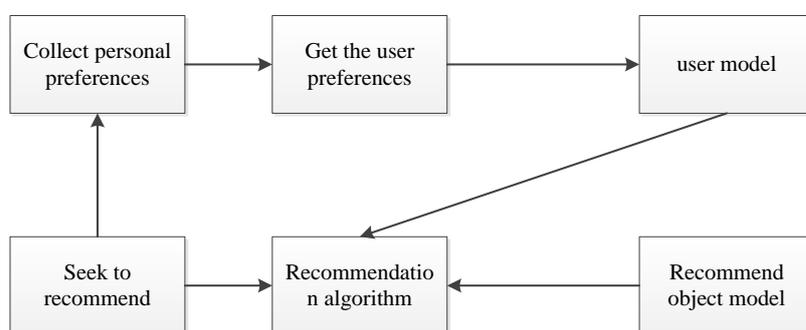
```
┌──────────────────┐      ┌──────────────────┐      ┌──────────────────┐
│ Collect personal │─────▶│   Get the user   │─────▶│                  │
│   preferences    │      │   preferences    │      │    user model    │
└──────────────────┘      └──────────────────┘      └──────────────────┘
         ▲                                                    │
         │                                                    ▼
┌──────────────────┐      ┌──────────────────┐      ┌──────────────────┐
│     Seek to      │─────▶│  Recommendatio   │◀─────│    Recommend     │
│    recommend     │      │   n algorithm    │      │   object model   │
└──────────────────┘      └──────────────────┘      └──────────────────┘
```

Figure 4: Recommender systems general model

5. Conclusion

Microblog as the most efficient, real-time information sharing platform, is favoured by users of all ages, whose value is not only to provide a platform for users to make speech and show themselves, but also a media platform to spread social focus and social values. With the development of all kinds of Weibo products, increasing number of the third-party developers, and the gradual improvement of mobile Internet infrastructure, etc., Weibo play a greater role in guiding public opinions. To study microblog products, customer relationship, and microblog content in its increasingly development period, has great practical significance.

The most successful recommendation algorithm existing now is the collaborative filtering recommendation algorithm which considered people's daily life habits, and divide into three models: projects collaborative recommendation, users collaborative and filtering collaborative recommendation.

The core of users collaborative recommendation is neighbours queries that can find the similarity between users, and the higher the similarity, the more similar the users are, then come to the information recommendation. Similar to users collaborative recommendation, projects collaborative recommendation is also based on the users' trust to a brand. Put it another way, if more people believe in on brand, there will be many users choose the brand of products. While recommendation based on model collaborative is somewhat different.

The difference is the machine learning and data applications on existing statistical methods for the testing model. Collaborative filtering algorithms have obvious advantages, which can be used for complex unstructured objects, help users discover new points of interest, and will gradually improve its performance with the increasing number of users. The disadvantage is that when the number of users increases, polyethylene and what they interested in are not the same, the comment of users may be different, and it cannot recommend the new user information. In addition, historical data has a great influence on the recommendation quality.

The thesis applies new algorithm method and uses topic detection and emotional orientation analysis method to analyse the microblog content. On the aspect of topic-finding, the author reserves microblog text terms and uses words credible association rules to co-occurrence and explore the field, and finally get great topic by the group consisting of nouns. As for the emotional propensity analysis, the article uses the word activation force

method to in-depth study relationship between the user's emotional words and topic terms and determine the relevant recommended level for a certain topic, so as to make information recommendation for centre users.

Acknowledgment

Reference

Bai H., Lin X., 2015, Sina weibo incident monitor and Chinese disaster microblogging classification, Journal of Digital Information Management, 13, 156-161, DOI: 10.1015/ 2015.04.013.

Guo Z., 2013, Unveiling the patterns of video tweeting: A sina weibo-based measurement study, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7799, 166-175, DOI: 10.1007/978-3-642-36516-4-17.

Hwang H.S., Choi E.K., 2016 ,Exploring gender differences in motivations for using Sina Weibo, KSII Transactions on Internet and Information Systems, 10, 1429-1441, DOI: 10.3837/tiis.2016.03.029.

Jia K., 2012, Detecting spam in Chinese microblogs - A study on Sina Weibo, Proceedings of the 2012 8th International Conference on Computational Intelligence and Security, 12, 578-581, DOI: 10.1109/CIS.2012.135.

Li D., 2016, Propagation regularity of hot topics in Sina Weibo based on SIR model - A simulation research, Proceedings - 2014 IEEE Computers, Communications and IT Applications Conference, 4, 310-315, DOI: 10.1109/ComComAp.2014.7017216.

Yu G., He F., 2016, An analysis of sleep complaints on Sina Weibo, Computers in Human Behavior, 62, 230-235, DOI: 10.1016/j.chb.2016.04.014.

Yuan B., Hu X., 2013, Understanding the top grass roots in Sina-Weibo, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7751, 17-24, DOI: 10.1007/978-3-642-36669-7_3.

Zhou Y., Zhou Q., Chen K., 2013, Observation of Matthew Effects in Sina Weibo microblogger, Proceedings - 2013 IEEE International Conference on Big Data, 3, 41-43, DOI: 10.1109/BigData.2013.6691796.